Fast Simulation for Slow Paths in Markov Models

Daniël Reijsbergen[1], Pieter-Tjerk de Boer[1], Werner Scheinhardt[1], and Boudewijn Haverkort[1,2]

[1]Center for Telematics & Information Technology, University of Twente, Enschede, The Netherlands
[2]Embedded Systems Institute, Eindhoven, The Netherlands
{d.p.reijsbergen, p.t.deboer, w.r.w.scheinhardt, b.r.h.m.haverkort}@utwente.nl

***Abstract*—Inspired by applications in the context of stochastic model checking, we are interested in using simulation for estimating the probability of reaching a specific state in a Markov chain *after* a large amount of time τ has passed. Since this is a rare event, we apply importance sampling. We derive approximate expressions for the sojourn times on a *given* path in a Markov chain conditional on the sum exceeding τ, and use those expressions to construct a change of measure. Numerical examples show that this change of measure performs very well, leading to high precision estimates in short simulation times.**

I. Introduction

Stochastic model checking is an increasingly important tool to support the design process of a variety of systems. The systems are modelled using a formalism like Petri nets, Markov reward models (MRMs), etc., and properties of these models are then verified [3]. Increasingly, these properties are stochastic in nature, and they often involve events that are hopefully rare, such as system failures.

Many methods for verifying such properties are known, but in the case of complex stochastic systems, statistical model checking using simulation is often the only feasible method. However, in order to efficiently simulate *rare* events, special techniques are needed. In a recent paper [5], an importance-sampling-based rare-event simulation method was developed for estimating probabilities of events of the form "absorption before a specific time" in a broad class of absorbing continuous-time Markov chains (CTMCs).

In the present work, we consider the opposite event, namely "absorption *after* a specific amount of time has passed by". While rarely studied in the rare event simulation literature, it is particularly motivated by MRMs where the event of interest is "collecting *sufficient reward* before absorption". Both problems can be shown to be equivalent [2].

For estimating such probabilities in general CTMCs, we envision a *two-step* approach: in the first step of each simulation run, the simulator samples a path (i.e., a sequence of states) through the chain, and in the second step it samples the sojourn times for that path. The present paper presents work in progress about the second subproblem: the probability of interest (i.e., that the sum of the sojourn times of a *given* path in a CTMC exceeds some threshold) is known in closed form [1], but its numerical evaluation is computationally expensive. Therefore, we derive an efficient importance sampling simulation algorithm for it, drawing sojourn times from a distribution that closely resembles the conditional distribution given the rare event of interest.

The rest of this paper is organised as follows. In Section II we study the conditional distribution of the sojourn times. In Section III we briefly introduce importance sampling simulation, and describe our algorithm. The good performance of our algorithm is illustrated experimentally in Section IV, and Section V provides some conclusions.

II. Conditional Sojourn Times

As noted above, we assume that a path ϕ through the Markov chain is already given, consisting of n states $x_1, \ldots, x_n$; only the sojourn times in the states on this path are unknown, but the rates of the states are given as $q_1, \ldots, q_n$, some of which may be identical. This path itself can be seen as an absorbing Markov chain on its own, as depicted in Figure 1. We now proceed to analyze the behaviour of the sojourn times T_j in the individual states j of this path, conditional on absorption occurring *after* some time bound τ. The results of this section will be used in Section III to obtain an efficient simulation algorithm.

The probability density of the sojourn time T_j is given by $f_j(x) = q_j e^{-q_j x}$, but we are interested in the distribution of T_j *conditional* on occurrence of the event $T > \tau$, where $T \triangleq \sum_{j=1}^{n} T_j$. Considering without loss of generality $j = 1$, we condition on the value of T_1 to find

$$\mathbb{P}(T_1 > t | T > \tau) = \int_t^\infty \frac{f_1(t_1)}{\mathbb{P}\left(T > \tau\right)} \mathbb{P}\left(T - T_1 > \tau - t_1\right) dt_1$$

and hence

$$f_1(t | T > \tau) = \begin{cases} \dfrac{f_1(t)}{\mathbb{P}\left(T > \tau\right)} \mathbb{P}(T - T_1 > \tau - t) & \text{if } t < \tau, \\ \dfrac{f_1(t)}{\mathbb{P}\left(T > \tau\right)} & \text{otherwise.} \end{cases} \tag{1}$$

This expression contains the probability $\mathbb{P}\left(T > \tau\right)$ which we are trying to estimate. Therefore our goal is now to obtain insight into the behaviour of $f_1(t|T > \tau)$ for large τ so we can construct a good approximation in the next section.

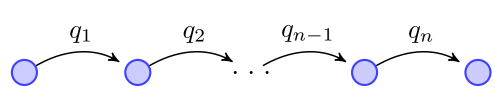

Figure 1. Path ϕ, seen as a Markov birth process.

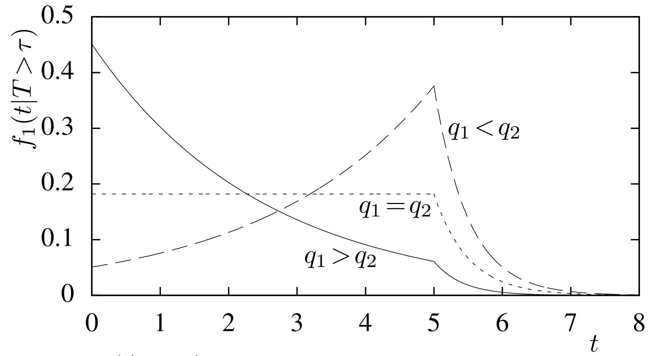

Figure 2. $f_1(t|T > \tau)$ for different parameter values q_1 and q_2, with $\tau = 5$. Solid line: $q_1 = 2.4, q_2 = 2$. Dotted line: $q_1 = 2.2, q_2 = 2.2$. Dashed line: $q_1 = 2, q_2 = 2.4$.

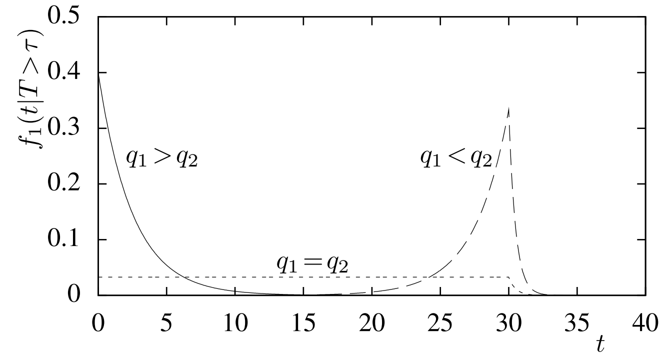

Figure 3. Same choices for the values of the parameters q_1 and q_2 as in Figure 2, but with $\tau = 30$.

We start by making (1) explicit for a two-state path $\phi = (x_1, x_2)$ with rates q_1 and q_2, $q_1 \neq q_2$. Then

$$f_1(t|T > \tau) = \begin{cases} \dfrac{q_1 e^{-(q_1-q_2)t}}{\frac{q_1}{q_1-q_2} + \frac{q_2}{q_2-q_1} e^{-(q_1-q_2)\tau}} & \text{if } t < \tau, \\ \dfrac{q_1 e^{-q_1 t}}{\frac{q_1}{q_1-q_2} e^{-q_2\tau} + \frac{q_2}{q_2-q_1} e^{-q_1\tau}} & \text{otherwise.} \end{cases} \tag{2}$$

Note that the same expression holds for $f_2(t|T > \tau)$ after interchanging q_1 and q_2.

The shape of this function for $t > \tau$ is always exponential with rate q_1. However, the shape of the part where $t < \tau$ depends on the parameter setting, where we distinguish three cases. For $q_1 > q_2$, this part is still *negative* exponential albeit with a different parameter, namely $q_1 - q_2$. However, for $q_1 < q_2$, this part is *positive* exponential, again with parameter $q_1 - q_2$. In between, as q_1 and q_2 become equal, this part approaches a constant. This can be seen in Figure 2.

In Figure 3 the time bound τ was increased sixfold, illustrating the limit behaviour of the system for large τ. For $q_1 > q_2$ we see that the probability mass right of τ vanishes, so we can approximate the function (2) by a simple exponential density with rate $q_1 - q_2$.

It is also interesting to observe how the *expected share* of the burden of consuming τ time units is distributed over the states. One easily derives the following from (2):

$$\mathbb{E}(T_1|T > \tau) \sim \begin{cases} \tau & \text{if } q_1 < q_2 \\ \tau/2 & \text{if } q_1 = q_2 \\ (q_2 - q_1)^{-1} & \text{if } q_1 > q_2, \end{cases}$$

with $\sim$ meaning that the ratio of left- and right-hand side goes to 1 as $\tau \to \infty$. This is illustrated in Figure 4. We see that when the rates differ, almost all of the time τ is typically spent in the state with the lowest rate, while the time spent in the other state tends to a constant.

These core observations do not just hold for two-state paths but for any path ϕ. Denote the lowest rate by β_1 and the second-lowest by β_2, and let r_i be the number of times rate β_i occurs on the path. Then, in the limit for large τ, a state i whose rate $q_i \neq \beta_1$ will contribute only an exponentially distributed amount of time with the *bounded* mean $(q_i - \beta_1)^{-1}$. If $r_1 = 1$, then the single state with rate β_1 will account for an amount that has an asymmetric Laplace distribution peaking at $t = \tau$ with rates $\beta_2 - \beta_1$ on the left side and β_1 on the right side. If there are $r_1 > 1$ states with rate β_1, then the expected contribution of each of these states is τ/r_1, and the conditional sojourn time in each state has an exponential distribution with rate β_1 to the right of τ, but a *polynomial* density with degree $r_1 - 2$ left of τ.

III. Simulation

We now proceed to construct an efficient simulation estimator for our probability of interest, namely $\mathbb{P}(T > \tau)$ given a path ϕ. The standard simulation estimator for this is

$$\hat{p} = \frac{1}{N} \sum_{i=1}^{N} \mathbf{1}_{\sum_j t_{ij} > \tau}, \tag{3}$$

where t_{ij} is the sampled sojourn time (with density $f_j(t)$) for state j in the ith simulation run, N is the number of simulation runs, and $\mathbf{1}$ the indicator function. This approach is very inefficient when the target event is rare. A remedy is *importance sampling* [4], where the samples t_{ij} are drawn from a different density $f_j^*(t)$ and weighted by a likelihood ratio:

$$\hat{p}^* = \frac{1}{N} \sum_{i=1}^{N} \prod_{j=1}^{n} \frac{f_j(t_{ij})}{f_j^*(t_{ij})} \mathbf{1}_{\sum_j t_{ij} > \tau}, \tag{4}$$

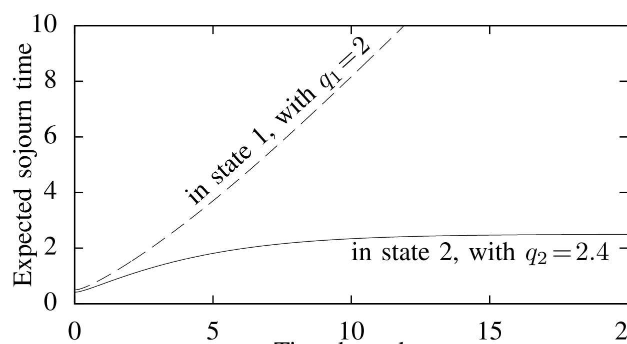

Figure 4. Expected sojourn times as a function of τ.

Since the exact calculation of $f_j(t|T > \tau)$ is problematic is general, we propose to use the following approximation instead, inspired by the findings in the previous section:

$$f_j^*(t) = \begin{cases} (q_j - \beta_1) \cdot e^{-(q_j - \beta_1) \cdot t} & \text{if } q_j > \beta_1 \\ r_1/\tau \cdot e^{-r_1/\tau \cdot t} & \text{if } q_j = \beta_1 \text{ and } r_1 = 1 \\ g(t|\beta_1, \beta_2) & \text{otherwise,} \end{cases} \tag{5}$$

with β_i and r_i defined as before, and $g(t|\beta_1, \beta_2)$ is given by the r.h.s. of (2) with each q_i replaced by β_i.

In practical applications where the Markov chain is not a pure-birth Markov process, the above algorithm for each simulation run i should be preceded by a phase in which the path itself (i.e., the set of states) is sampled, possibly also using importance sampling (cf. the two-phase approach discussed in Section I).

IV. Simulation Results

In this section we empirically demonstrate the effectiveness of the method. Throughout this section, all results are based on 10^6 simulation runs. We compare standard Monte Carlo (MC) simulation using (3) to our importance sampling (IS) approach using (4) and (5). In the first two examples, we also give the true values for p, computed directly using the Erlang and hypoexponential distribution functions.

In Table I, we consider a two-state path ϕ with unequal rates. We see that the method works well; the very slow increase of the relative error (r.e.) (defined as 1.96 times the sample standard deviation of the estimator $\hat{p}$ (or $\hat{p}^*$) divided by the sample mean $\hat{p}$ (or $\hat{p}^*$) itself) as τ becomes bigger, suggests the relative error is in fact upper-bounded.

τ	$\hat{p}$	MC-r.e.	$\hat{p}^*$	IS-r.e.	true
5	2.52E-4	0.1235	2.417E-4	0.0047	2.417E-4
7	8.0E-6	0.6929	4.71E-6	0.0054	4.736E-6
9	0	—	8.947E-8	0.0060	8.93E-8
100	0	—	8.3E-89	0.0078	8.3E-89

Table I
Simulation results, $q_1 = 2$, $q_2 = 2.4$.

τ	$\hat{p}$	MC-r.e.	$\hat{p}^*$	IS-r.e.	true
5	2.03E-4	0.1375	2.011E-4	0.0058	2.004E-4
7	3.0E-6	1.1316	3.372E-6	0.0070	3.363E-6
100	0	—	6.29E-94	0.0279	6.3E-94

Table II
Simulation results, $q_1 = q_2 = 2.2$.

In Table II, we set $q_1 = q_2$. The results are still good but somewhat less accurate, which can be explained by the poor resemblance between $f_1^*(t)$ and $f_1(t|T > \tau)$ for $t < \tau$.

Finally, in Table III we show results for a path with 50 states and 25 different rates. Note that direct calculation of the true probability is bot numerically feasible in this case. Having demonstrated that the method works well for the pure-birth processes for which it was intended (as the second step of the two-step approach), we also give an example *derived from* a Markov-reward model involving an M/M/5/5 queue. The resulting Markov chain is a birth-death process with 6 states labeled $0, \ldots, 5$, with birth rates $0.1 \cdot \exp(5 - k)$ in state $k = 0, \ldots, 4$, and death rates $10 \cdot k \cdot \exp(5 - k)$ in state $k = 1, \ldots, 5$. We start in 5 and the absorbing state is 0.

τ	$\hat{p}$	MC-r.e.	$\hat{p}^*$	IS-r.e.	true
12	2.092E-2	0.0134	2.097E-2	0.0051	—
20	1.4E-5	0.5238	1.727E-5	0.0070	—
100	0	—	2.19E-39	0.0180	—

Table III
Simulation results, $q_i = \lceil \frac{i+1}{2} \rceil$, $i = 1, \ldots, 50$.

Generating appropriate sample paths using standard simulation, and then drawing sojourn times with the algorithm from the present paper, leads to the results reported in Table IV. These results look promising, with a relative error growing just linearly in τ. For larger τ however, generating sample paths becomes more difficult, and importance sampling will be needed here as well.

τ	$\hat{p}$	MC-r.e.	$\hat{p}^*$	IS-r.e.	true
1	2.315E-2	0.0127	2.294E-2	0.0068	—
2	1.48E-4	0.1611	1.776E-4	0.0176	—
4	0	—	9.895E-9	0.0563	—

Table IV
Simulation results for the M/M/5/5 queue.

V. Conclusion and Future Work

We have found explicit results and useful approximations for the conditional distribution of sojourn times on a given path in a Markov chain, given that their sum exceeds a bound. The resulting expressions are relatively simple and yield insight into how this rare event typically happens. Based on these insights we have constructed an importance sampling change of measure and shown that performs well. Future research will first focus on more general Markov chains, where we need to identify the most probable paths leading to the rare event, after which we can apply the method presented here to those paths. Also it will be interesting to consider (rare) events in which both a time- and reward-bound play a role.

References

[1] S.V. Amari and R.B. Misra. Closed-Form Expressions for Distribution of Sum of Exponential Random Variables. *IEEE Transactions on Reliability*, 46(4):519–522, 1997.

[2] C. Baier, L. Cloth, B.R. Haverkort, H. Hermanns, and J.P. Katoen. Performability assessment by model checking of Markov reward models. *Formal Methods in System Design*, 36(1):1–36, 2010.

[3] C. Baier and J.P. Katoen. Principles of model checking. 2008.

[4] P. Heidelberger. Fast simulation of rare events in queueing and reliability models. *Performance Evaluation of Computer and Communication Systems*, pages 165–202, 1993.

[5] D. Reijsbergen, P.T. de Boer, W. Scheinhardt, and B. Haverkort. Rare event simulation for highly dependable systems with fast repairs. In *Seventh International Conference on the Quantitative Evaluation of Systems*, pages 251–260. IEEE, 2010.